Exercise 1: Bayesian Linear Model

In the Bayesian linear model, we assume that the data follows the following law:

$$y = f(\mathbf{x}) + \epsilon = \boldsymbol{\theta}^T \mathbf{x} + \epsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and independent of $\mathbf{x}$. On the data-level this corresponds to

$$y^{(i)} \quad = \quad f\left(\mathbf{x}^{(i)}\right) + \epsilon^{(i)} = \boldsymbol{\theta}^T \mathbf{x}^{(i)} + \epsilon^{(i)}, \quad \text{for } i \in \{1, \ldots, n\}$$

where $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$ are iid and all independent of the $\mathbf{x}^{(i)}$'s. In the Bayesian perspective it is assumed that the parameter vector $\boldsymbol{\theta}$ is stochastic and follows a distribution.

Assume we are interested in the so-called maximum a posteriori estimate of $\boldsymbol{\theta}$, which is defined by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{y}).$$

(a) Show that if we choose a uniform distribution over the parameter vectors $\boldsymbol{\theta}$ as the prior belief, i.e.,

$$q(\boldsymbol{\theta}) \propto 1,$$

then the maximum a posteriori estimate coincides with the empirical risk minimizer for the L2-loss (over the linear models).

(b) Show that if we choose a Gaussian distribution over the parameter vectors $\boldsymbol{\theta}$ as the prior belief, i.e.,

$$q(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2\tau^2} \boldsymbol{\theta}^\top \boldsymbol{\theta}\right], \qquad \tau > 0,$$

then the maximum a posteriori estimate coincides for a specific choice of τ with the regularized empirical risk minimizer for the L2-loss with L2 penalty (over the linear models), i.e., the Ridge regression.

(c) Show that if we choose a Laplace distribution over the parameter vectors $\boldsymbol{\theta}$ as the prior belief, i.e.,

$$q(\boldsymbol{\theta}) \propto \exp\left[-\frac{\sum_{i=1}^{p} |\boldsymbol{\theta}_i|}{\tau}\right], \qquad \tau > 0,$$

then the maximum a posteriori estimate coincides for a specific choice of τ with the regularized empirical risk minimizer for the L2-loss with L1 penalty (over the linear models), i.e., the Lasso regression.

Exercise 2: Covariance Functions

Consider the commonly used covariance functions mentioned in the lecture slides: constant, linear, polynomial, squared exponential, Matérn, exponential covariance functions.

(a) Show that they are valid covariance functions. Note that the proofs for the Matérn and exponential covariance functions are out of scope for this exercise. Additionally, you may use the following composition rules. In these rules we assume that $k_0(\cdot, \cdot)$ and $k_1(\cdot, \cdot)$ are valid covariance functions:

(i) $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$ is a valid covariance function;
(ii) $k(\mathbf{x}, \mathbf{x}') = c \cdot k_0(\mathbf{x}, \mathbf{x}')$ is a valid covariance function if $c \geq 0$ is constant;
(iii) $k(\mathbf{x}, \mathbf{x}') = k_0(\mathbf{x}, \mathbf{x}') + k_1(\mathbf{x}, \mathbf{x}')$ is a valid covariance function;
(iv) $k(\mathbf{x}, \mathbf{x}') = k_0(\mathbf{x}, \mathbf{x}') \cdot k_1(\mathbf{x}, \mathbf{x}')$ is a valid covariance function;
(v) $k(\mathbf{x}, \mathbf{x}') = g(k_0(\mathbf{x}, \mathbf{x}'))$ is a valid covariance function if g is a polynomial function with positive coefficients;
(vi) $k(\mathbf{x}, \mathbf{x}') = t(\mathbf{x}) \cdot k_0(\mathbf{x}, \mathbf{x}') \cdot t(\mathbf{x}')$ is a valid covariance function, where t is any function;
(vii) $k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}'))$ is a valid covariance function;
(viii) $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{A} \mathbf{x}'$ is a valid covariance function if $\mathbf{A} \succeq 0$.

(b) Are these covariance functions stationary or isotropic? Justify your answer.